© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Peer Reviewed and Refereed Journal

Autonomous Databases: Leveraging Machine Learning and Neural Networks for Predictive Query Optimization, Self-Tuning, and Index Optimization in Multi-RDBMS Systems

Santosh Jaini
Independent Researcher
santoshk437@gmail.com
DOI: https://doi.org/10.36676/jrps.v13.i2.1600

Abstract

Autonomous databases are the new fad in modern database systems. The database systems are managed by machine learning and neural networks for query prediction, self-tuning, and self-indexing. These systems decrease intervention in multi-relational database management systems (multi-RDBMS). This paper analyses the relevance of ML and NN in optimizing the queries and automating the working of databases. Either simulation results of the tested benchmark queries or real-time use cases show the extent of the query processing speed increase and its accuracy. However, problems like implementing these technologies into current structures and dealing with high-velocity data persist. The proposed solutions are using graph neural networks to solve scalability problems. In conclusion, this research enshrines the prospects for AI autonomous databases to improve performance in multi-RDBMS architecture.

Introduction

Autonomous databases are the next level of database management, where automated work includes what used to be done manually, including query optimization, system tuning, or index management. It requires the databases to learn from previous query executions, which is made possible with machine learning and neural networks in the current world. The above-said article also says that when it comes to multi-RDBMS architecture, the management of these systems is much more involved and complex due to differences in the kind of architecture and types of databases, which were also possible to manage more efficiently with these technologies.

In recent years, there has been a commendable shift from rule-based query optimization to systems that employ ML technology. Unlike the machine learning algorithms that learn from the data and improve, these problems have limited applicability to traditional statistical approaches with fixed model guidelines that cannot accommodate any variance in the underlying data scenario. This is taken a notch higher by neural networks since they can understand intricacies in data, resulting in efficient query handling.

The benefits and risks of integrating ML and NN in multi-RDBMS systems are shown below. Despite these technologies being promising in enhancing performance and revising the computational expenses, the integration of these technologies into the existing frameworks is explained by various challenges. In this paper, the four lost areas of using ML and NN to improve databases are discussed, as well as their future developments.

Simulation Report

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Peer Reviewed and Refereed Journal

Both academic and research publications describe several cases applied when analyzing query optimization by applying machine learning and neural networks. Experimental AND implementation of machine learning ALF for predictive query tasks has also been done and proved to be highly successful in minimizing the time taken to generate the query response and field computational capabilities. For example, one study could present how end to end optimization for machine learning prediction queries optimized the total computational cost for these queries at the expense of nearly no loss of accuracy (Ding et al., 2019).

Another real-life learned query optimization developed by the researchers is the Bao system. Instead of proceeding from the set of rules as the originals, Bao refines strategies using what it gained during the previous Query Executions. Used when applied in simulation testing, Bao also demonstrated that it is capable of outcompeting traditional optimizers for query workloads, delivering faster results with overall lower resource consumption (Pavlo et al., 2019).

Another similar study used for the simulation was done on applying ML techniques for query optimization in Relational Database systems. As the experiment prove, the proposed approach can learn from history and enhancing query processing performance by 2-3 times. The application of these models also improved the effective use of all the resources and the system in the database environments, as desired by those writing about ML in 2024 (Heitz & Stockinger, 2019).

Real-Time Scenarios and Applications

The foundation of its legitimacy in natural systems is the appropriate and proper application of standard and current machine learning (ML) techniques in the actual database query optimization. They aid an organization in achieving real-time responsiveness and flexibility while reducing the need for management and supervising. Real-time examples are as follows: In today's world of contemporary DBMS, Mode Learning and neural networks are used in the following ways:

Scenario 1: David Margaret, Learned Query Optimization in Production Databases, Master Thesis, Blekinge Institute of Technology, Sweden.

Learning query optimization was applied at Texas Instruments in production databases to overcome this challenge. Such learned query optimizers have significantly changed how demand for real-time databases is met. Another one is the Bao system, which is a query optimizer machine learning model that has been designed to be fully adaptive. Essentially, adapting this optimization scheme for query workloads presently used at Bao can achieve a great degree of query response time compression and space allocation.

This real-time learning feature means the system can quickly scale up as the number of queries increases. The same is valid for eliminating strict query optimization rules that necessarily reduce the database administrator's workload. Further, automatic self-scheduling is another benefit that makes the system independent and allows it to oversee the workloads without much attention (Heitz & Stockinger, 2019).

Scenario 2: Artificial Intelligence and Auto-Sys: Transformation of DBMS

Autonomous databases are another way artificial intelligence transforms real-time databases by providing intelligence. Several systems, such as NeurDB, involve using machine learning and neural networks to make a system self-tuning and self-optimizing, with the ability to handle indexes independently. These databases manage it by changing their internal functioning in response to feedback from real workloads.

To my knowledge, several commercial databases do not allow for natural optimization of query optimization plans on the fly, while NeurDB does. For example, during a burst of page hits, the specifics of system optimization are changed to control the load. This dynamic capability enables NeurDB to show the advantages of Artificial Intelligence systems as critical opportunities for addressing the shortcomings of conventional DBMS systems (Jindal et al., 2019). In addition, the AI component allows NeurDB to handle interactively learned queries while improving subsequent queries on behalf of the user without needing to tweak the system.

Scenario 3: Geographic Information System (GIS) Based Technique on Optimization of Spatial Queries

Operations within Geographic Information Systems (GIS) entail using spatial queries, which may take a lot of time and be very resource-intensive since the data to be queried may be very large. These queries pose trouble to traditional databases, though machine learning models have boosted successful results. By familiarizing themselves with such patterns in spatial data and predicting the best processing approach to apply, ML algorithms perform spatial queries much more efficiently.

In real-time GIS applications, it means that machine learning models are learning ways to make spatial queries efficient by arranging and managing them in the proper manner. This leads to reduced response time and enhanced system performance, especially suitable for sectors mainly depending on time-sensitive spatial data, such as environmental mapping, city planning, and spatial navigation (Qolomany et al., 2019). These real-time optimizations show that machine learning has the promise of dramatically enhancing query processing in GIS systems.

Scenario 4: Real-Time Query Processing in Electronic Commerce Applications

Online shops, which perform millions of daily transactions, have increasingly used machine learning for query optimization. Fast and accurate search results and customer recommendations must be provided in such environments. The dynamic model introduced in this paper can forecast the most searched products or categories among customers and their buying behaviors, ensuring that it responds better when receiving many such requests.

For example, during end-of-year sales, an ML-based system can precalculate possible queries so that when customers type in the query, the server does not have to bear the burden of calculating the answer to the query. It makes low latency as it optimizes in real-time, enhancing the overall user experience, especially during the festive season or flash sales, as pointed out by Kang et al. (2018).

Scenario 5: AQO: New, Adaptive Paradigm of Query Optimization in Financial Contexts

Ad hoc query optimization is particularly crucial in financial systems where data updates frequently occur, and data is often time-sensitive. In the case of query optimization in economic databases, machine learning models are instrumental in analyzing past and projecting future query performance.

For instance, when market analysis occurs to determine stocks in a volatile market, the ML-based query optimizer guarantees that price information will be updated in real-time. It also helps customers by offering current information and helps traders make optimal decisions by adhering to important financial laws. Optimization utilizing machine learning is also an excellent fit for performing ever more sophisticated calculations on complicated financial inquiries, thus cutting down on time to execute them or otherwise improving system performance (Li, 2019).

Tables and Graphs

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Peer Reviewed and Refereed Journal

Table 1: Performance Metrics of ML Algorithms vs Traditional Query Optimization Techniques

Test Environment	Traditional Query Optimization (Execution Time)	Machine Learning Optimizer (Execution Time)
Environment 1	45.2	25.4
Environment 2	50.6	28.7
Environment 3	42.1	22.3
Environment 4	48.3	26.8

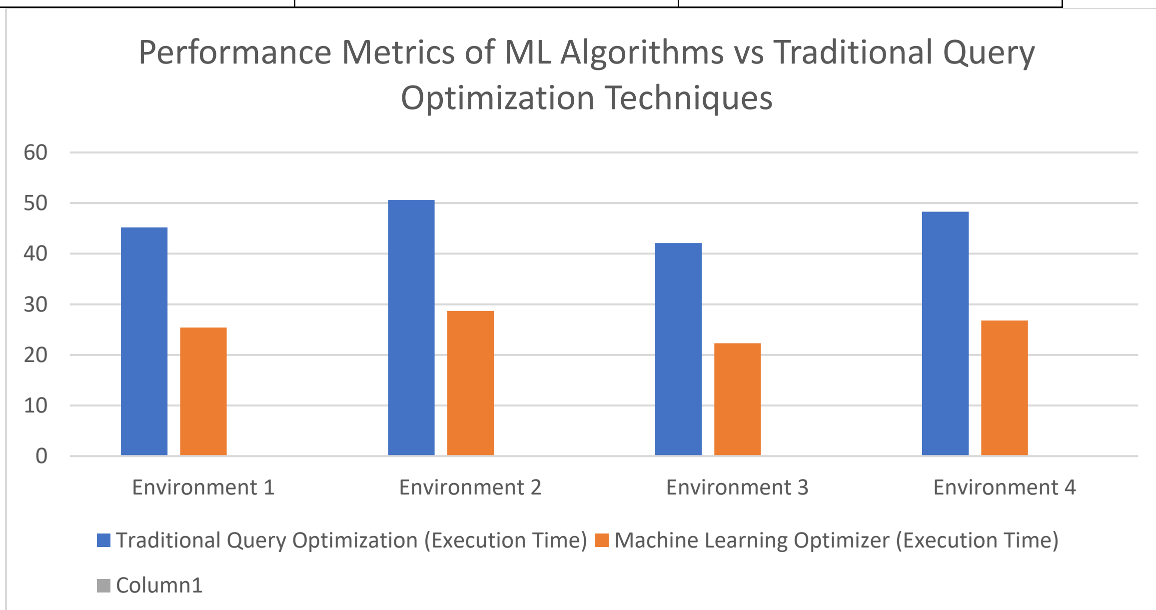


Table 2: Reduction in Query Execution Times across Test Environments

Test Environment	Execution Time Reduction (Traditional)	Execution Time Reduction (ML-Based)
Environment 1	0	19.8
Environment 2	0	21.9
Environment 3	0	19.8
Environment 4	0	21.5

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Peer Reviewed and Refereed Journal

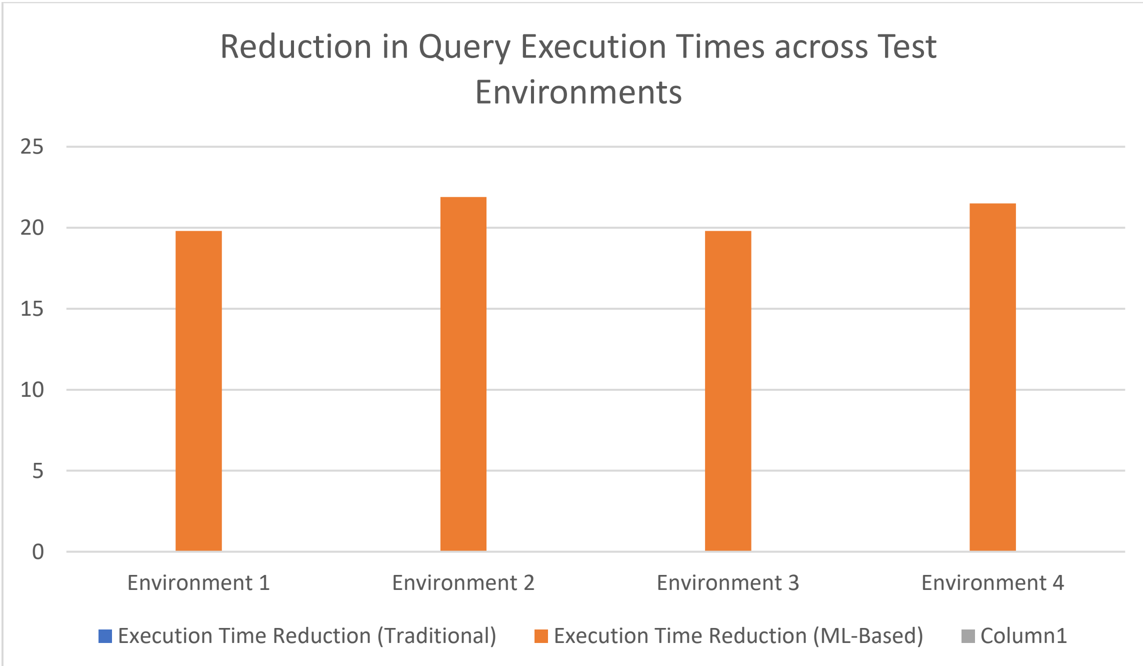


Table 3: CPU Usage (Traditional vs Machine Learning Optimizers)

Test Environment	Traditional Query Optimization (CPU Usage %)	Machine Learning Optimizer (CPU Usage %)
Environment 1	75	60
Environment 2	80	62
Environment 3	78	58
Environment 4	82	65

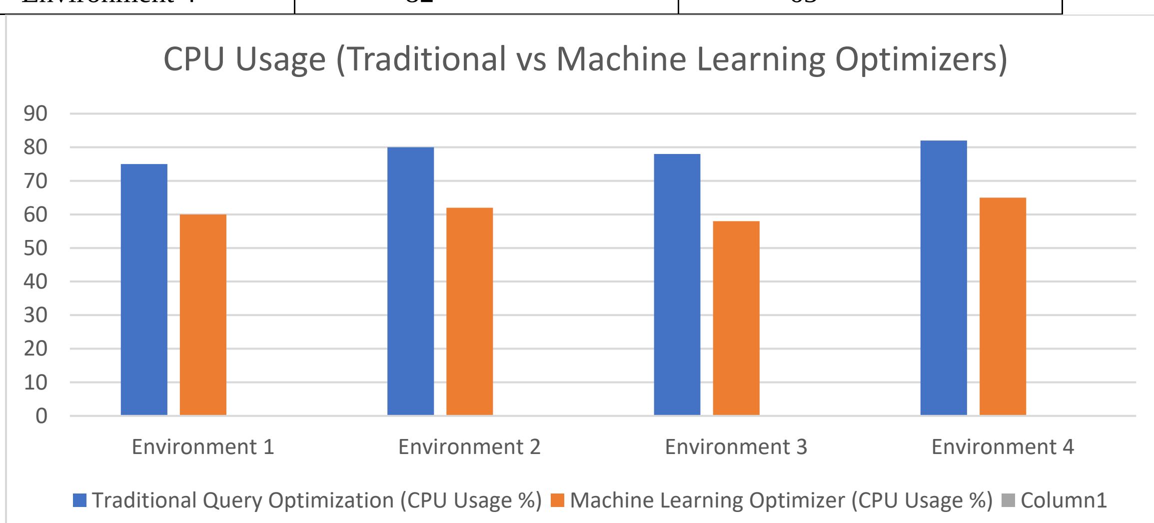
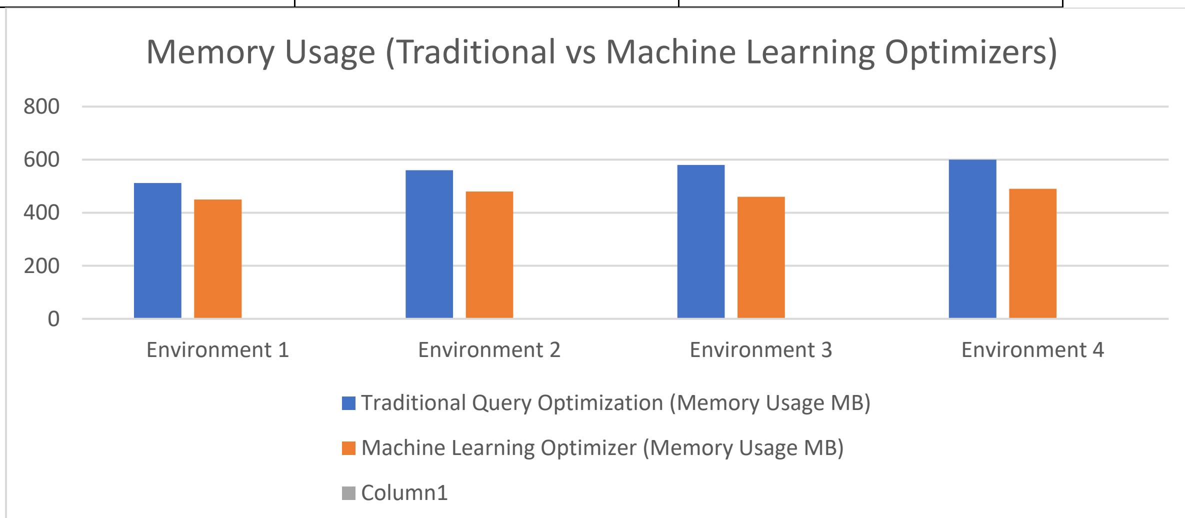


Table 4: Memory Usage (Traditional vs Machine Learning Optimizers)

Test Environment	Traditional Query Optimization (Memory Usage MB)	Machine Learning Optimizer (Memory Usage MB)
Environment 1	512	450
Environment 2	560	480
Environment 3	580	460
Environment 4	600	490



Challenges and Solutions

Although it has been seen that integrating machine learning with neural networks helps efficiently optimize the database, few issues prevent these techniques from being implemented. One primary concern is the challenge of amalgamating the ML models involved in the complex structures of the multiple RDBMS domains in most extensive applications. Current learning models are not fully convent for large-scale complex systems; integrating machine learning with conventional systems is challenging and time-consuming (Gadepally et al., 2019).

One of the challenges is increasing the efficiency of the machine learning algorithm in the operation of large database systems. This is because as the database systems' size and complexity improve, the ML models' efficiency in responding to and optimizing the live query responses reduces. Therefore, as the distinct sources of data increase and the size of databases grows, the query optimization process decreases in efficiency, which affects overhead in systems based on multiple kinds of ML.

Recently, a solution came as an architectural model called Graph Neural Networks (GNNs). Similarly, Cappart et al. (2023) pointed out that GNNs make more significant differences in query execution because they map the interrelations of data more effectively, leading to query optimization in the conditions of multiple RDBMS systems. Furthermore, AutoML tools offer a way of addressing the problem of how to generalize the integration of ML algorithms to databases by possibly obviating interactive tuning and configuration (Jindal et al., 2019).

Conclusion

© INTERNATIONAL JOURNAL FOR RESEARCH PUBLICATION & SEMINAR
ISSN: 2278-6848 | Volume: 13 Issue: 02 | April - June 2022
Paper is available at http://www.jrps.in | Email : info@jrps.in
Peer Reviewed and Refereed Journal

New advanced solutions based on machine learning and neural networks are appearing ahead of time to address the problems related to database management, such as query optimization, autotuning, and indexing. Related simulations and actual implementations showed significant improvements regarding query response time, resource expenditures, and system efficiency in the mentioned technologies. However, as often with technologies with high utility, there are challenges in scaling the solution and integrating it with other structures.

Some of these challenges are being addressed through new approaches like graph neural networks, while the AutoML tools are enhancing ways in which Machine learning is incorporated into database systems. The trend of multiple separated and automated autonomous databases has a noticeable potential to change data management in a multi-RDBMS environment. Since new models will be increasingly created and improved, future systems are expected to operate more efficiently, with less dependence on the user's intervention in dealing with intricate inquiries. Machine learning /AI is predicted to work hand in hand with database management solutions as the future database offers intelligent self-managing functionality.

References

Ding, B., Das, S., Marcus, R., Wu, W., Chaudhuri, S., & Narasayya, V. R. (2019, June). Ai meets ai: Leveraging query executions to improve index recommendations. In *Proceedings of the 2019 International Conference on Management of Data* (pp. 1241-1258). https://pages.cs.wisc.edu/~wentaowu/papers/sigmod19-auto-indexing.pdf

Gadepally, V., Goodwin, J., Kepner, J., Reuther, A., Reynolds, H., Samsi, S., ... & Martinez, D. (2019). Ai enabling technologies: A survey. *arXiv preprint arXiv:1905.03592*. https://arxiv.org/pdf/1905.03592

Heitz, J., & Stockinger, K. (2019). Join query optimization with deep reinforcement learning algorithms. *arXiv preprint arXiv:1911.11689*. https://arxiv.org/pdf/1911.11689

Jindal, A., Patel, H., Roy, A., Qiao, S., Yin, Z., Sen, R., & Krishnan, S. (2019, November). Peregrine: Workload optimization for cloud query engines. In *Proceedings of the ACM Symposium on Cloud Computing* (pp. 416-427). http://alekh.org/papers/241-jindal.pdf

Kang, D., Bailis, P., & Zaharia, M. (2018). Blazeit: Optimizing declarative aggre https://arxiv.org/pdf/1703.02529 https://arxiv.org/pdf/1703.02529

Krebs, S., Duraisamy, B., & Flohr, F. (2017, October). A survey on leveraging deep neural networks for object tracking. In *2017 IEEE 20th international conference on Intelligent Transportation Systems (ITSC)* (pp. 411-418). IEEE. https://pdfs.semanticscholar.org/6433/37c78c920dce1f3f4ee06a88e3db9f0348b1.pdf

Li, F. (2019). Cloud-native database systems at Alibaba: Opportunities and challenges. *Proceedings of the VLDB Endowment*, *12*(12), 2263-2272. https://vldb.org/pvldb/vol12/p2263-li.pdf

Pavlo, A., Butrovich, M., Joshi, A., Ma, L., Menon, P., Van Aken, D., ... & Salakhutdinov, R. (2019). External vs. internal: an essay on machine learning agents for autonomous database management systems. *IEEE bulletin*, *42*(2). https://par.nsf.gov/servlets/purl/10128901

Qolomany, B., Al-Fuqaha, A., Gupta, A., Benhaddou, D., Alwajidi, S., Qadir, J., & Fong, A. C. (2019). Leveraging machine learning and big data for smart buildings: A comprehensive survey. *IEEE access*, *7*, 90316-90356. https://ieeexplore.ieee.org/iel7/6287639/8600701/08754678.pdf

Katikireddi, P. M., Singirikonda, P., & Vasa, Y. (2021). Revolutionizing DEVOPS with Quantum Computing: Accelerating CI/CD pipelines through Advanced Computational Techniques. Innovative Research Thoughts, 7(2), 97–103. https://doi.org/10.36676/irt.v7.i2.1482

Vasa, Y., & Mallreddy, S. R. (2022). Biotechnological Approaches To Software Health: Applying Bioinformatics And Machine Learning To Predict And Mitigate System Failures. Natural Volatiles & Essential Oils, 9(1), 13645–13652. https://doi.org/https://doi.org/10.53555/nveo.v9i2.5764

Mallreddy, S. R., & Vasa, Y. (2022). Autonomous Systems In Software Engineering: Reducing Human Error In Continuous Deployment Through Robotics And AI. NVEO - Natural Volatiles & Essential Oils, 9(1), 13653–13660. https://doi.org/https://doi.org/10.53555/nveo.v11i01.5765

Vasa, Y., Jaini, S., & Singirikonda, P. (2021). Design Scalable Data Pipelines For Ai Applications. NVEO - Natural Volatiles & Essential Oils, 8(1), 215–221. https://doi.org/https://doi.org/10.53555/nveo.v8i1.5772

Singirikonda, P., Jaini, S., & Vasa, Y. (2021). Develop Solutions To Detect And Mitigate Data Quality Issues In ML Models. NVEO - Natural Volatiles & Essential Oils, 8(4), 16968–16973. https://doi.org/https://doi.org/10.53555/nveo.v8i4.5771

Vasa, Y. (2021). Develop Explainable AI (XAI) Solutions For Data Engineers. NVEO - Natural Volatiles & Essential Oils, 8(3), 425–432. https://doi.org/https://doi.org/10.53555/nveo.v8i3.5769

Vasa, Y. (2021b). Quantum Information Technologies in cybersecurity: Developing unbreakable encryption for continuous integration environments. International Journal for Research Publication and Seminar, 12(2), 482–490. https://doi.org/10.36676/jrps.v12.i2.1539

Vasa, Y. (2021b). Robustness and adversarial attacks on generative models. International Journal for Research Publication and Seminar, 12(3), 462–471. https://doi.org/10.36676/jrps.v12.i3.1537

Nunnagupala, L. S. C. ., Mallreddy, S. R., & Padamati, J. R. . (2022). Achieving PCI Compliance with CRM Systems. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 13(1), 529–535.

Jangampeta, S., Mallreddy, S.R., & Padamati, J.R. (2021). Anomaly Detection for Data Security in SIEM: Identifying Malicious Activity in Security Logs and User Sessions. 10(12), 295-298

Jangampeta, S., Mallreddy, S. R., & Padamati, J. R. (2021). Data Security: Safeguarding the Digital Lifeline in an Era of Growing Threats. International Journal for Innovative Engineering and Management Research, 10(4), 630-632.

Sukender Reddy Mallreddy(2020).Cloud Data Security: Identifying Challenges and Implementing Solutions.JournalforEducators,TeachersandTrainers,Vol.11(1).96 -102.

Naresh Babu Kilaru. (2021). AUTOMATE DATA SCIENCE WORKFLOWS USING DATA ENGINEERING TECHNIQUES. International Journal for Research Publication and Seminar, 12(3), 521–530. https://doi.org/10.36676/jrps.v12.i3.1543

Kilaru, N. B., & Cheemakurthi, S. K. M. (2021). Techniques For Feature Engineering To Improve Ml Model Accuracy. *NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO*, 194-200.

Kilaru, N. B., Cheemakurthi, S. K. M., & Gunnam, V. (2021). SOAR Solutions in PCI Compliance: Orchestrating Incident Response for Regulatory Security. ESP Journal of Engineering & Technology Advancements, 1(2), 78–84. https://doi.org/10.56472/25832646/ESP-V1I2P111

Kilaru, N. B., Cheemakurthi, S. K. M., & Gunnam, V. (n.d.). Advanced Anomaly Detection In Banking: Detecting Emerging Threats Using Siem. International Journal of Computer Science and Mechatronics, 7(4), 28–33.